PATENT ABSTRACTS OF JAPAN

(11)Publication number :   2002-202790

(43)Date of publication of application :   19.07.2002

---

(51)Int.Cl.   G10L 13/06
                                           G10L 13/08
                                           G10L 13/00

---

(21)Application number : 2000-401041   (71)Applicant :   YAMAHA CORP

(22)Date of filing :   28.12.2000   (72)Inventor :   KENMOCHI HIDENORI
                                                            XAVIER SERA
                                                            JORDI BONADA

---

(54) **SINGING SYNTHESIZER**

(57)Abstract:
PROBLEM TO BE SOLVED: To synthesize singing voices of high quality.
SOLUTION: A spectrum model synthesis(SMS), which is an analytical and synthetic
process, is conducted about the phoneme or two or more phoneme chains, a database
10 is prepared, and the SMS data of the phoneme or phoneme chains required for
synthesis are concatenated and synthesized, to obtain singing voices. Into the database
10, separate segment data are stored by each different pitch, dynamics, and tempo
concerning the same phoneme or phoneme chain. A harmonic component adjustment
means 22 and a non-harmonic component adjustment means 23 adjust the harmonic
components and non-harmonic components of read segment data so as to match them
to a target pitch. A duration adjustment means 24 adjusts the length of the phonemes
or the phoneme chains with the length matching the target tempo. A segment level
adjustment means 25 carries out level adjustment, and then connects individual
segments, generates harmonic components corresponding to the desired pitch, and
synthesizes high quality singing voices with a non-harmonic components and the
harmonic component.
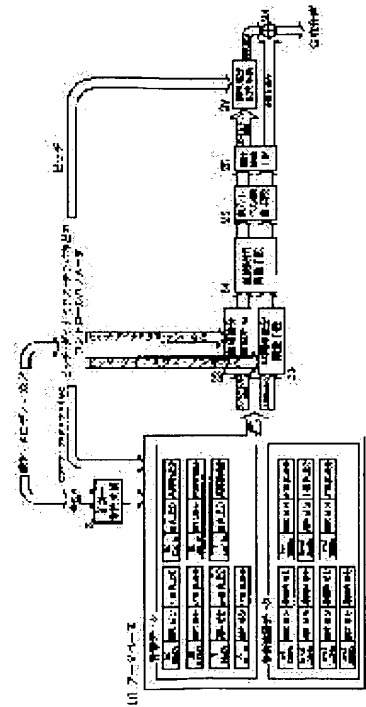
LEGAL STATUS

[Date of request for examination]   16.10.2001
[Date of sending the examiner's decision of rejection]   17.08.2004
[Kind of final disposal of application other than the examiner's
decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]   2004-19147
[Date of requesting appeal against examiner's decision of rejection] 16.09.2004
[Date of extinction of right]
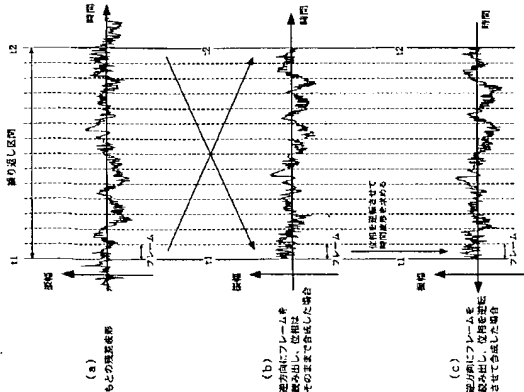
DRAWINGS
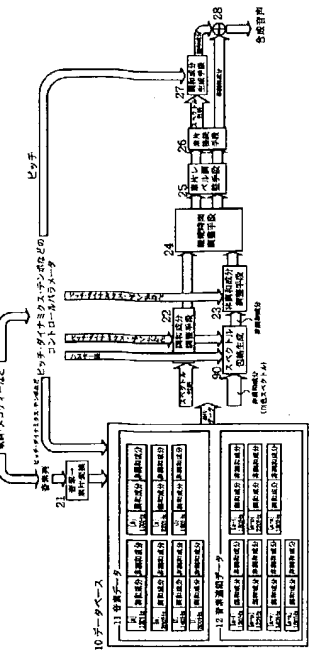
[Drawing 1]

[Drawing 3]

[Drawing 10]

[Drawing 13]

[Drawing 2]

[Drawing 4]

[Drawing 5]

[Drawing 6]

[Drawing 14]

[Drawing 7]

[Drawing 11]

[Drawing 8]

[Drawing 15]

[Drawing 16]

[Drawing 9]

[Drawing 12]

[Drawing 17]
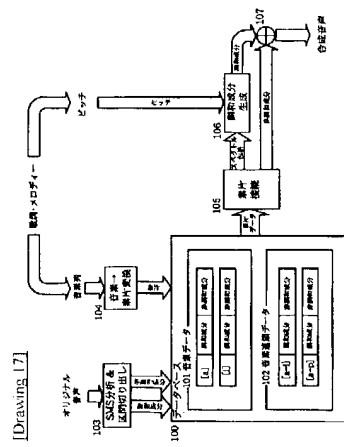
[Translation done]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is drawing for explaining creation processing of the phoneme database used for the song synthesizer unit of this invention
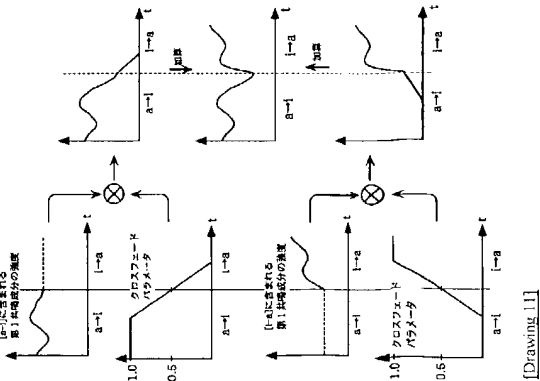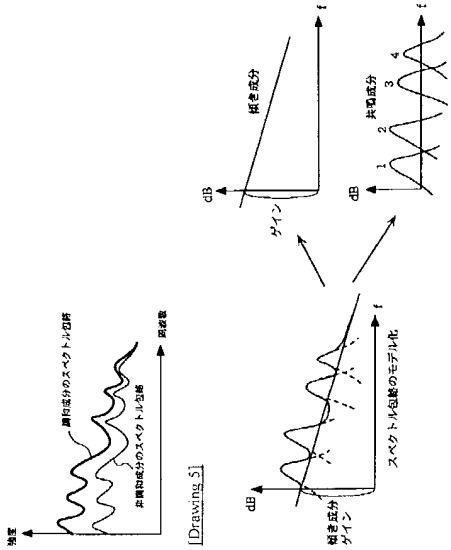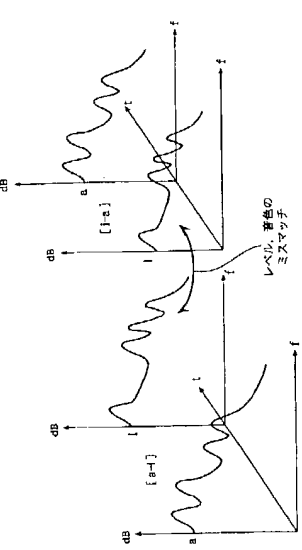
[Drawing 2] It is drawing for explaining the song sound composition processing in the song synthesizer unit of this invention.
[Drawing 3] It is drawing for explaining the non-harmonic component adjustment processing in the song synthesizer unit of this invention.
[Drawing 4] It is drawing for explaining the loop-formation processing in the song synthesizer unit of this invention.
[Drawing 5] It is drawing for explaining modeling of a spectral envelope.
[Drawing 6] base -- it is drawing for explaining the mismatch in the connection of piece data.
[Drawing 7] It is drawing for explaining the smoothing processing in the song synthesizer unit of this invention.
[Drawing 8] It is drawing for explaining the level adjustment processing in the song synthesizer unit of this invention.
[Drawing 9] It is the functional block diagram showing the configuration of the gestalt of 1 operation of the song synthesizer unit of this invention in a detail.
[Drawing 10] It is drawing showing an example of the hardware for operating the song synthesizer unit of this invention.
[Drawing 11] It is drawing in which lengthening and showing an example of the spectral envelope of the harmonic component in a sound, and a non-harmonic component.
[Drawing 12] It is drawing for explaining creation processing of the phoneme database in the gestalt of other operations of the song synthesizer unit of this invention.
[Drawing 13] It is drawing showing the example of 1 configuration of a spectrum whitening means.
[Drawing 14] It is drawing for explaining the song sound composition processing in the gestalt of other operations of the song synthesizer unit of this invention.
[Drawing 15] It is drawing for explaining control whenever husky.
[Drawing 16] It is drawing showing the example of a configuration of the spectral envelope generation means at the time of enabling it to perform control whenever husky.
[Drawing 17] It is drawing for explaining the song synthesizer unit which applied the conventional SMS method.
[Description of Notations]
10 a Phoneme Database and 13 An SMS Analysis Means and 14 A Section Logging Means and 21 Phoneme -> -- Base -- Piece Conversion Means and 22 Harmonic-Component Adjustment Means and 23 Non-harmonic Component Adjustment Means and 24 Duration Adjustment Means and 25 Base -- Piece Level Adjustment Means and 26 Base -- Piece Connecting Means and 27 Harmonic-Component Generation Means and 28 Synthetic Means and 80 Spectrum Whitening Means and 90 Spectral Envelope Generation Means

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the song synthesizer unit which compounds singing voice.
[0002]
[Description of the Prior Art] Conventionally, the attempt which is going to compound singing voice has been performed broadly. It is application of one regulation speech synthesis of them, and the pitch data and words data corresponding to a musical interval of a note are considered as an input, and it compounds using the regulation speech synthesizer for text-to-speech syntheses. In many cases, that which analyzed and parameterized the raw data point or it which makes a unit the phoneme chain containing a phoneme (or a phoneme: phoneme) or two phonemes or more is accumulated in a database, and a voice element (a phoneme or phoneme chain) required at the time of composition is chosen, and it connects and compounds. For example, please refer to JP,62-6299,A, JP,10-124082,A, JP,11-1184490,A, etc. However, since these techniques aimed at talking originally and compounding language, when compounding singing voice, quality was not what can not necessarily be satisfied.
[0003] For example, when delicate fluctuation of a vibrato indispensable to the song voice to which the part which is lengthening the sound which influences the quality of a song sound most becomes unnatural in many cases in the wave superposition composite system represented by PSOLA (Pitch-Synchronous OverLap and Add) although the intelligibility of a synthetic song sound is good, or a pitch was performed, there was a trouble of becoming unnatural composite tone in many cases. if it is going to compound song voice using the wave connection mold speech synthesizer of the large-scale corpus base, since [moreover,] the wave of a basis will be connected and outputted, without processing it at all in principle -- the base of an astronomical figure -- piece data are needed.
[0004] On the other hand, the synthetic vessel aiming at composition of singing voice is also devised from the start. For example, the composite system by the characteristic-frequency-region composite system is known (JP,3-200300,A). Although this lengthens and the quality and the vibrato of a sound, and the degree of freedom of pitch change are large, the articulation of composite tone (especially a consonant part) is low, and quality cannot necessarily be satisfied.
[0005] By the way, as shown in a U.S. Pat. No. 5029509 specification, the technique called spectrum modeling composition (SMS:Spectral Modeling Synthesis) of performing analysis and composition of musical sound using the model which expresses an original sound with two components (deterministic component), i.e., a harmonic component, and a non-harmonic component (stochastic component) is known. While the musical description of musical sound is controllable good, in the case of singing voice, it is expectable according to this SMS analysis composition, that articulation also with the amount of high child Otobe is obtained with use of a non-harmonic component. Therefore, if this technique is applied to composition of singing voice, it is expected that the composite tone having high articulation and musicality will be acquired. Although the proposal of the concrete application about the technique of compounding a sound based on an SMS analysis composition technique is actually performed by patent No. 2906970, the methodology in the case of using an SMS technique for song composition (singing synthesizer) at coincidence is also described.
[0006] The song synthesizer unit which applied the technique proposed by said patent No. 2906970 is explained with reference to drawing 17 . In drawing 17 , the phoneme database 100 is created by carrying out SMS analysis, starting input voice to every voice element (a phoneme or phoneme chain), and memorizing it in SMS analysis and the section logging section 103. The voice element data in a database 100 (the phoneme data 101, phoneme chain data 102) consist of data of the single put in order by time series or two or more frame trains, and temporal responses, such as the SMS data corresponding to each frame, i.e., the spectral envelope of a harmonic component, a spectral envelope of a non-harmonic component, and a phase spectrum, are memorized. the phoneme train which constitutes desired words when compounding a song sound -- asking -- phoneme -> -- base -- by the piece transducer 104, a voice element (a phoneme or phoneme chain) required to constitute the phoneme train is determined, and the SMS data (a harmonic component and non-harmonic component) of a required voice element are read from said database 100. and base -- the harmonic-overtone component which has a desired pitch is generated, connecting serially the SMS data of the voice element read in the piece connection 105, and maintaining the configuration of the spectral envelope according to the pitch information corresponding to the melody of the musical piece in the harmonic-component generation section 106 about a harmonic component. for example, in compounding with "SAITA" (saita) the base [#s], [s], [s-a], [a], [a-i], [i], [i-t], [t], [t-a], [a], and [a#] -- a piece -- connecting -- base -- the harmonic component of a desired pitch is generated, with the configuration of the spectral envelope contained in the SMS data obtained by connection of a piece maintained. And synthesized speech is obtained by adding this harmonic component and non-harmonic component that were generated with the synthetic means 107, and changing into the data of a time domain.
[0007]
[Problem(s) to be Solved by the Invention] Thus, by using an SMS technique, intelligibility is good and becomes possible [obtaining a synthetic song sound natural also about the lengthened part]. However, if the method stated by the above-mentioned patent No. 2906970 is too primitive and simple and singing voice is compounded with the method, the following troubles will produce it.
- Since the configuration of the spectral envelope of the harmonic component of a voiced sound changes with pitches a little, when compounding in a different pitch from the time of analysis, a good tone is not obtained if it remains as it is.

- In order that a harmonic component may remain in a remainder component slightly even if it removes a harmonic component in the case of a voiced sound when performing SMS analysis, if it compounds to the song sound of a pitch which is different from the original sound, using the remainder component (non-harmonic component) same as mentioned above as it is, a remainder component will float and it will become a cause of ******** at ********* and a noise.
- Since the phoneme data as an analysis result of SMS and phoneme chain data are piled up in time as it is, adjustment of the time amount which lengthens a sound, or the time amount of the change between phonemes cannot be performed. That is, it cannot be made to sing by desired Il Tempo.
- It is easy to generate a noise at the time of connection of a phoneme or a phoneme chain.
[0008] Then, this invention materializes the technique in the case of using for song composition the SMS technique proposed in the above-mentioned patent No. 2906970, adds the still more extensive amelioration about the quality of composite tone, and aims at offering the song synthesizer unit which solved each above-mentioned trouble. Moreover, while being able to make size of said database small, it aims at offering the song synthesizer unit which raised the effectiveness of database creation. Furthermore, it aims at offering the song synthesizer unit which can adjust the husky degree of synthesized speech.
[0009]
[Means for Solving the Problem] In order to attain the above-mentioned purpose, the song synthesizer unit of this invention It has the phoneme database which memorized the data of a harmonic component, and the data of a non-harmonic component about the voice element which is the phoneme chain which is relation of a phoneme or two phonemes or more. By reading the voice element data corresponding to words from said phoneme database, and connecting A duration adjustment means to adjust the time amount length of the voice element data which are the song synthesizer unit which compounds a song sound, and were read from said phoneme database so that target Il Tempo and how to sing might be suited, It has an adjustment means to adjust said harmonic component and said non-harmonic component of the voice element data read from said phoneme database so that the target pitch might be suited. moreover, the base which performs smoothing processing or level adjustment processing about a harmonic component and each non-harmonic component when connecting said voice element data -- it has a piece level adjustment means. Furthermore, in said phoneme database, a pitch, dynamics, and two or more voice element data with which Il Tempo differs are memorized about the same phoneme or the phoneme chain. In said phoneme database, the voice element data which consist of the voice element data which consist of a phoneme chain from a vowel or a vowel to a consonant, voice element data which consist of a phoneme chain from a consonant to a consonant, and a phoneme chain from a vowel to a vowel are memorized further again from the voice element data which a vowel etc. lengthens and consist of a sound, and a consonant.
[0010] further -- again -- the data of said harmonic component, and the data of said non-harmonic component -- the base -- it memorizes as a data stream of the frequency domain corresponding to each frame of the frame train included at the section of a piece. Said duration adjustment means generates the frame train of desired time amount length further again repeating 1 in the frame train included in a voice element, or two or more frames, or by thinning out a frame. Said duration adjustment means reverses the phase of the phase spectrum of the non-harmonic component further again, when repeating the frame of a non-harmonic component and it goes back in time at the time of composition. It has a harmonic-component generation means to change only a pitch into a desired pitch, maintaining the facies of the spectral envelope of the harmonic component contained in voice element data about a harmonic component further again at the time of song sound composition.
[0011] It lengthened further again among the voice element data memorized in said phoneme database, and the flat spectrum obtained by carrying out the multiplication of the inverse number of the spectrum which lengthens and represents the section of a sound to the magnitude spectrum of the non-harmonic component is memorized as a magnitude spectrum of a non-harmonic component about the voice element corresponding to a sound. The magnitude spectrum of a non-harmonic component is obtained by lengthening, calculating the magnitude spectrum of a non-harmonic component based on the magnitude spectrum of the harmonic component, and multiplying said flat spectrum by it about the non-harmonic component of a sound, further again at the time of song sound composition. The part in said phoneme database lengthens, said flat spectrum which the magnitude spectrum of the non-harmonic component is not memorized, but others lengthen, and is memorized by the voice element of a sound about the voice element about a sound is used further again, and i is the thing which lengthens and compounds a sound. When calculating the magnitude spectrum of a non-harmonic component based on the magnitude spectrum of said harmonic component, according to the parameter which controls whenever [husky], the gain in 0Hz of the magnitude spectrum of said non-harmonic component to calculate is controlled further again.
[0012] It lengthens at the time of song sound composition, and it is the thing which is lengthened and is used as a magnitude spectrum of the non-harmonic component of a sound further again about the magnitude spectrum obtained by the magnitude spectrum of the non-harmonic component of a sound by carrying out the multiplication of the inverse number of the representation [lengthen and] magnitude spectrum within the sound section, creating a flat spectrum, calculating the magnitude spectrum according to the parameter which lengthens and controls whenever [husky] based on the magnitude spectrum of the harmonic component of a sound, and multiplying by this magnitude spectrum and said created flat spectrum.
[0013]
[Embodiment of the Invention] The song synthesizer unit of this invention carries out [voice / input] SMS analysis, asks for the SMS data of a harmonic component and a non-harmonic component, and has the phoneme database which started the required section and was packed for every phoneme and every phoneme chain. In this database, the information which shows the music expression of the information which shows the pitch of that voice element as a header in addition to the information on a phoneme or a phoneme chain and dynamics, Il Tempo, etc. is also included. the sensuous information whether the voice element (a phoneme or phoneme chain) is the sound of Town & Country, and whether dynamics information is the sound of mezzo forte here -- you may be -- or the base -- you may be the physical information which shows the level of a piece. Moreover, it has an SMS analysis means to decompose into a non-harmonic component and a harmonic component, and to analyze input song voice for said database creation. Moreover, it has a means (automatic and hand control are not asked) for starting the phoneme or phoneme chain (base piece) to need.
[0014] The example of said phoneme database creation is explained with reference to drawing 1 . In drawing 1 , 10 is a phoneme database, like the phoneme database 100 mentioned above, SMS analysis of the input song voice is carried out by the SMS analyzor 13,

and the SMS data (the base SMS data of each frame contained in the piece) for every piece of ** started by every phoneme or phoneme chain (voice element) by the section logging section 14 are stored. however, this phoneme database 10 -- setting -- base -- it memorizes as separate data for every pitch from which piece data differ, different dynamics, and different II Tempo.

[0015] In addition, in making Japanese words sing, a voice element consists of vowel data (one frame or two or more frames), the data (multiple frame) from a consonant to a vowel or the data (multiple frame) from a vowel to a consonant, data (multiple frame) from a consonant to a consonant, and data (multiple frame) from a vowel to a vowel. Although considered as the unit which usually records VCV (consonant a vowel, - vowel) or CVC (consonant - vowel, a consonant) longer than syllable etc. on a phoneme database in voice synthesizers, such as regulation composition In the song synthesizer unit of this invention especially aiming at composition of a song sound The vowel which may set to a song and appears is pronounced for a long time, and is developed, and the data of a vowel (valve flow coefficient) or a vowel to a consonant (VC), the data of a consonant to a consonant, and the data of a vowel to a vowel are stored in a phoneme database from the data of a sound, and a consonant.

[0016] Said SMS analyzor 13 carries out [voice / original / input song] SMS analysis, and outputs the SMS analytical data for every frame. Namely, input voice is divided into a series of time frames, and frequency analysis is carried out by FFT etc. for every frame. It asks for a magnitude spectrum and a phase spectrum, and the spectrum of the specific frequency corresponding to the peak of a magnitude spectrum is extracted from the frequency spectrum (complex spectrum) obtained as a result as a line spectrum. Let a spectrum with the frequency near fundamental frequency and the frequency of that integral multiple be a line spectrum at this time. This extracted line spectrum supports said harmonic component. And a remainder spectrum is obtained by subtracting the line spectrum extracted as mentioned above from the spectrum of an input wave of the frame. Or the time amount data point of the harmonic component compounded from said extracted line spectrum is subtracted from the input data point of the frame, the time amount data point of a remainder component is obtained, and a remainder spectrum is obtained by carrying out frequency analysis of this. Thus, the obtained remainder spectrum corresponds to said non-harmonic component (SUTOKASU tick component).

[0017] In addition, the frame period used for said SMS analysis may be a fixed fixed length, or may be a variable-length period which changes the period according to the pitch of input voice etc. What is necessary is to process input voice with the 1st fixed-length frame period, to detect the pitch, and just to adopt the technique of changing the period of the frame which follows with the pitch which reworked or obtained input voice from the analysis result of the frame in front of the frame with the frame period according to the result, in making a frame period into variable length.

[0018] In the section logging section 14, the SMS analytical data outputted for every frame from said SMS analyzor 13 are started so that it may correspond to the die length of the voice element memorized in a phoneme database. That is, the phoneme chain of a vowel phoneme, a vowel, a consonant or a consonant, and a vowel, the phoneme chain of a consonant and a consonant, and the phoneme chain of a vowel and a vowel are started hand control or automatically so that it may be most suitable for composition of a song sound. Here, the data (lengthening sound) of the long section which is developing and singing the vowel as a vowel phoneme are also cut down. Moreover, in this section logging section 14, the pitch of that input voice is detected from said SMS analysis result. this pitch detection - that base -- it asks for an average pitch from the frequency of the line spectrum of a low degree of the harmonic components of the frame contained in a piece, and is carried out by averaging this about all frames.

[0019] Thus, the data of the harmonic component and the data of a non-harmonic component are cut down for every piece of **, information, such as a pitch of the input song voice, dynamics showing a music expression, and II Tempo, is further added as a header, and it stores in said phoneme database 10. An example of the phoneme database 10 created by doing in this way is shown in drawing 1 , and the phoneme data area 11 corresponding to a phoneme and the phoneme chain data area 12 corresponding to a phoneme chain are shown in the phoneme database 10. And signs that a vowel [a] lengthens in said phoneme data area 11, four kinds of phoneme data, the pitch frequency of 130Hz, 150Hz, 200Hz, and 220Hz, and a vowel [i] lengthen to it to a sound, and three kinds of phoneme data, the pitch frequency of 140Hz, 180Hz, and 300Hz, are stored in it to the sound are shown. As opposed to the phoneme chain [a-i] which shows relation of a phoneme [a] and [i] to said phoneme chain data area 12 Moreover, the pitch frequency of 130Hz, and two 150Hz kinds, Signs that each 100Hz phoneme chain data is stored to 140Hz, 180Hz, and a phoneme chain [a-z] to two kinds and a phoneme chain [a-s] (120Hz and 220Hz) to the phoneme chain [a-p] are shown. In addition, although the case where the data with which pitches differ to the same phoneme or a phoneme chain are stored is shown, it memorizes as different data similarly as mentioned above here about the data with which the music expressions of the dynamics of the input song voice, II Tempo, etc. differ.

[0020] in addition, each base -- the data showing the harmonic component contained in piece data, and a non-harmonic component About the SMS data from said SMS analyzor 13 started by said section logging section 14 for every piece of **, i.e., a harmonic component the base -- all the spectral envelopes (the reinforcement (amplitude) of a line spectrum (harmonic-overtone sequence) and spectrum of a phase) of each frame contained in a piece are memorized as they are -- Or you may memorize by the approach of ******** memorized not as the spectral envelope itself but as a thing which expressed the spectral envelope with a certain function. Or a harmonic component may be memorized in the form of a time amount wave which carried out inverse transformation. moreover -- a non-harmonic component -- the base -- you may memorize as the spectrum (magnitude spectrum) on the strength and phase spectrum of each frame of the section corresponding to a piece, and may memorize in the form of the time amount data point of the section itself. moreover, each above-mentioned storage format -- immobilization -- it is not necessary to be -- base -- you may make it make the storage formats differ according to the properties (for example, a nasal, a fricative, an explosive sound, etc.) of the voice of every piece and its section In addition, by the following explanation, the data of said harmonic component are memorized in the form of a spectral envelope and a non-harmonic component is explained as what is memorized in the form of the magnitude spectrum and a phase spectrum. In the case of such a storage format, storage capacity needed can be lessened. Thus, two or more data corresponding to the music expression of a different pitch to the same phoneme or the same phoneme or dynamics, II Tempo, etc. are stored in the phoneme database 10 in the song synthesizer unit of this invention.

[0021] Next, the synthetic processing of a song sound using the phoneme database 10 created in this way is explained with reference to drawing 2 . In drawing 2 , 10 is the phoneme database mentioned above. 21 [moreover,] -- phoneme -> -- base -- the base for searching said phoneme database 10 for the phoneme train corresponding to the words data of the musical piece which is a piece conversion means and should compound a song sound -- it changes into a piece. For example, it is ****** [s] to the input of the phoneme train of

"second_a_i_t_a". [s-a] [a] [a-i] [i] [i-t] [t] [t-a] [a] is outputted. the base read from said phoneme database 10 based on control parameters with which 22 is contained in the melody data of said musical piece etc., such as a pitch, dynamics, and Il Tempo, -- a harmonic-component adjustment means to adjust the data of the harmonic component of the piece data, and 23 are non-harmonic component adjustment means to adjust to the data of said non-harmonic component. 24 -- the base from said harmonic-component adjustment means 22 and said non-harmonic component adjustment means 23 -- a duration adjustment means to change the duration of piece data -- the base to which 25 adjusts level of each piece data of ** from said duration adjustment means 24 -- a piece level adjustment means -- 26 -- said base -- the base which connects to time series each piece data of ** by which level adjustment was carried out with the piece level adjustment means 25 -- a piece connecting means -- 27 -- said base -- the base connected by the piece connecting means 26 -- a harmonic-component generation means to generate the harmonic component (harmonic-overtone component) of a desired pitch based on the data (spectral envelope information) of the harmonic component of the piece data -- the harmonic-overtone component by which 28 was generated with said harmonic-component generation means 27, and said base -- it is an addition means to compound the non-harmonic component outputted from the piece connecting means 26. Synthesized speech is obtained by changing the output of this addition means 28 into the signal of a time domain.

[0022] Hereafter, the processing in each above-mentioned block is explained to a detail. said phoneme -> -- base -- the piece conversion means 21 generates ****** from the phoneme train changed based on input words, and, thereby, chooses the voice element in the phoneme database 10 (a phoneme and phoneme chain). as mentioned above, even if it is the same phoneme and a phoneme chain, corresponding to a pitch, dynamics, Il Tempo, etc., two or more things (voice element data) store in a database -- having -- **** -- base -- according to various control parameters, the optimal thing is chosen at the time of piece selection. Moreover, it does not choose, but some candidates are chosen and you may make it ask for the SMS data used for composition with those interpolation. The harmonic component and non-harmonic component as a result of SMS analysis are stored in the selected voice element. As for these contents, SMS data (reinforcement and phase), i.e., the spectral envelope of a harmonic component, the spectral envelope (reinforcement and phase) of a non-harmonic component, or the wave itself is contained. Among these, based on **, a harmonic component and a non-harmonic component are generated so that a desired pitch and the duration demanded may be suited. For example, it asks for the spectral envelope of harmony and a non-harmonic component with interpolation etc., or a spectrum configuration is made to deform so that a desired pitch may be suited.

[0023] Adjustment processing of a harmonic component is performed with the [adjustment of harmonic component] aforementioned harmonic-component adjustment means 22. In the case of the voiced sound, about the harmonic component, the reinforcement of the harmonic component which it is as a result of SMS analysis, and the spectral envelope of a phase are contained. base -- when a piece is plurality, or it chooses the thing optimal out of it for desired control parameters (pitch etc.) -- or two or more base -- it asks for the spectral envelope which was suitable for the desired control parameter out of the piece with actuation of interpolation etc. Moreover, the obtained spectral envelope may be made to transform by a certain approach corresponding to still more nearly another control parameter Moreover, in order to make the sound which becomes jarring mitigate or to give the description to a sound, a filter which passes only a fixed band may be covered. In addition, in the case of non-vocal sound, there is no harmonic component.

[0024] Since the effect of the original pitch remains in the non-harmonic component of the SMS analysis result of [adjustment of non-harmonic component] voiced sound, a sound may become unnatural when compounding the sound of another pitch. In order to prevent this, it is necessary to the low-pass component of a non-harmonic component to perform actuation which suits a desired pitch. This actuation is performed with said non-harmonic component adjustment means 23. With reference to drawing 3 , the adjustment actuation to this non-harmonic component is explained. (a) of drawing 3 is the example of the magnitude spectrum of the non-harmonic component obtained when SMS analysis is carried out [sound / voiced]. As shown in this drawing, it is difficult to remove the effect of a harmonic component completely, and some crest is made in near the harmonic overtone. When voice is compounded in a pitch other than the pitch of a basis, using this non-harmonic component as it is, the mountains near a low-pass harmonic overtone are perceived, and a ******** case is in a jarring sound, without melting into a harmonic component well. Then, although what is necessary is just to change the frequency of a non-harmonic component according to change of a pitch, since there is little effect of a harmonic component from the first, it is desirable [the non-harmonic component of a high region] to use a magnitude spectrum from the first as it is. That is, what is necessary is just to perform compression and expanding of a frequency shaft according to the pitch for which it asks in low-pass. However, don't change the original tone at this time. That is, it is necessary to perform this processing, with the facies of a magnitude spectrum maintained.

[0025] (b) of drawing 3 is drawing showing the result of having performed above-mentioned processing. As shown in this drawing, three low-pass crests are moved to the right according to the desired pitch. Spacing of the crest of a mid-range is narrowed and the crest of a high region remains as it is. Height is adjusted so that each crest may maintain the facies of the magnitude spectrum shown with a broken line. In addition, since there is no effect of the original pitch in the case of non-vocal sound, the above-mentioned actuation is unnecessary. Moreover, corresponding to a control parameter, some actuation (for example, deformation of a spectral envelope configuration etc.) may be further performed to the obtained non-harmonic component. Moreover, in order to make the sound which becomes jarring mitigate or to give the description to a sound, a filter which passes only a fixed band may be covered.

[0026] [duration adjustment], now as [this] -- base -- since the die length from the first which a piece has will be used as it is, singing voice is compoundable only to fixed timing. then -- if required according to the timing for which it asks -- base -- it is necessary to change the continuation length of a piece the case of for example, a phoneme chain -- base -- thinning out the frame contained in a piece -- base -- the die length of a piece becomes short and it can lengthen by making it overlap. moreover -- for example, -- the case (when it lengthens and is a sound) where the number of phonemes is one -- base -- if only the frame in a piece part is used -- lengthening -- a part -- short -- becoming -- base -- it can lengthen by repeating the inside of a piece.

[0027] lengthening -- the case of a sound -- base -- when repeating the inside of a piece, rather than it repeats only an one direction -- an one direction -- progressing -- hard flow -- return -- it progresses in the direction of origin again (that is, the loop formation of within the fixed section or the random section is carried out) -- ** -- the way to repeat what is said Although it is known that the noise of a knot will be mitigable, when a non-harmonic component is divided into every frame (immobilization or variable length) and is memorized in the frequency domain, it is a problem to repeat the frame data of a frequency domain in a form as it is, and to compound a wave. This is

because it must be made for the wave in a frame itself to have to become reverse in time when going to hard flow in time. What is necessary is to reverse the phase of a frequency domain and just to change into a time domain, in order to generate the wave which goes to hard flow in time from the frame data of the original frequency domain. Drawing 4 is drawing showing this situation.

[0028] (a) of drawing 4 is drawing showing the wave of a non-harmonic component from the first. It shall repeat and develop progressing to t2 from the repeat section t1 shown in drawing, going to hard flow in time, after amounting to t2, and going to the forward direction, after amounting to t1 again, and the non-harmonic component for a sound shall be generated. As mentioned above, a non-harmonic component is divided for every immobilization or variable-length frame, and is memorized by the frequency component. What is necessary is to carry out reverse FFT of the frame data of a frequency domain, and just to compound, hanging a windowing function and making it overlap, in order to generate the wave of a time domain. It becomes the wave from which only the sequence of a frame became reverse in time [the wave in a frame] with origin as it was shown in drawing 4 (b), when having read a frame into hard flow in time here, having compounded and the frame data of a frequency domain were changed into a time domain as it is, it becomes discontinuity, and becomes causes, such as a noise and distortion.

[0029] What is necessary is just to process frame data beforehand so that a reverse wave may be generated in time in case the wave of a time domain is searched for from frame data in order to solve this. Since it is g(t) =f (- t) and f (t) and g (t) are real variable functions when the wave of a basis is set to f (t) (the wave following infinity is considered for convenience), the wave which becomes hard flow in time is set to g (t) and each Fourier transform is set to F (omega) and G (omega), it is G(omega) =F(omega) * (* shows a complex conjugate).

It ******. Since a complex conjugate becomes what made the phase reverse when expressed with the amplitude and a phase, in order to generate a reverse wave in time, it turns out that what is necessary is just to make all the phase spectrums of the frame data of a frequency domain into reverse. If it does in this way, as shown in (c) of drawing 4 , the interior of a frame will also serve as a reverse wave in time, and neither a noise nor distortion will produce it.

[0030] the base above with said duration adjustment means 24 -- compression processing (infanticide of a frame) of a piece, expanding processing (repeat of a frame), and loop-formation processing (in the case [Lengthening.] of a sound) are performed. Thereby, the duration (namely, frame queue length) of each read piece of ** can be adjusted to the desired die length.

[0031] [-- base -- piece level adjustment] -- further -- base -- a piece and base -- when a difference is in the configuration of the spectral envelope of each component of harmony and not harmonizing, too much in the connection part of a piece, there is ******** fear as a noise. This is cancelable by carrying out smoothing of the connection part, applying two or more frames. This smoothing processing is explained with reference to drawing 5 - drawing 7 . a non-harmonic component -- base -- even if dispersion in a tone or level is in the connection of a piece -- comparatively -- ****** -- being hard -- a sake -- here -- a harmonic component -- smoothing -- it shall carry out Data are made easy to treat at this time, and in order to simplify count, suppose that the spectral envelope of a harmonic component is divided into the resonance component which was expressed with the straight line or the exponential function and which it inclined and was expressed with the component, the exponential function, etc., and is considered as shown in drawing 5 . Here, the reinforcement of a resonance component shall be calculated on the basis of an inclination component, shall add an inclination component and a resonance component, and shall express a spectral envelope. That is, the function showing the spectral envelope [harmonic component] using said inclination component and resonance component is expressing. Here, suppose that it inclines and the value which extended said inclination component to 0Hz is called the gain of a component.

[0032] two base as shown in drawing 6 at this time -- [i-a] shall be connected with a piece [a-i] Since a mismatch is in the tone and level of i of a connection in order to collect from another sound recording from the first, as it is shown in drawing 6 , a level difference wave-like in a connection part occurs, and each piece of ** is ******** as a noise. then -- a core [connection / the] -- carrying out -- order -- a number of frames -- applying -- each base -- if cross fade of each parameter of an inclination component and a resonance component contained in a piece is carried out, the level difference in a connection part disappears and generating of a noise can be prevented. What is necessary is to multiply each parameter of the piece of both ** by function (cross fade parameter) which is set to 0.5 in a connection part, and just to add it, as shown in drawing 7 in order to carry out cross fade of each parameter. The example shown in drawing 7 shows the example which carries out cross fade of this to the motion in each piece of ** of [a-i] of the reinforcement (based on the inclination component) of the 1st resonance component, and [i-a]. thus, the thing for which the multiplication of the cross fade parameter is carried out to each parameter (in this case, each resonance component), and it is added to it -- base -- generating of the noise in the connection of a piece can be prevented.

[0033] moreover, instead of carrying out cross fade as mentioned above -- base -- the amplitude of order becomes almost the same in the connection part of a piece -- as -- base -- level adjustment of each component of harmony and not harmonizing may be performed. [of a piece] level adjustment -- base -- it can carry out by applying a strange multiplier at regularity or the time to the amplitude of a piece. The case where [a-i] and [i-a] are connected and compounded is taken for an example like the above, and it explains per example of level adjustment. Here, it considers doubling the gain of the inclination component of each of said piece of **. As shown in (a) of drawing 8 , and (b), it asks for difference with the gain of an actual inclination component first on the basis of what carried out linear interpolation of the gain of the inclination component between the first frame and last frame (broken line in drawing) about each piece of ** of [a-i] and [i-a]. Next, it asks for the typical sample (each parameter of an inclination component and a resonance component) of each phoneme of [a] and [i]. The data of the frame of the beginning of [a-i] and the last frame may be used for this. If the difference based on this representation sample for which it subsequently asked by asking for a linear interpolation thing [parameter] first in the top is added, as it is shown in (c) of drawing 8 , since all parameters surely become the same, the discontinuity of the gain of an inclination component is not generated on a boundary. About other parameters, such as a parameter of a resonance component, discontinuity can be prevented similarly. In addition, it is not based on the approach described above, for example, the data of a harmonic component are changed into a data point, and it may be made to perform level adjustment etc. in a time domain.

[0034] said base -- the base mentioned above in the piece level adjustment means 26 -- the base after smoothing between pieces or level adjustment processing was performed -- the piece connecting means 26 -- base -- piece connection processing is performed. And in the harmonic-component generation means 27, by generating the harmonic-overtone train corresponding to a desired pitch, with the obtained harmonic-component spectral envelope maintained, an actual harmonic component is obtained and a synthetic song sound is

obtained by adding a non-harmonic component to it. And this is changed into the signal of a time domain. For example, when it has both the components of harmony and not harmonizing, by the frequency component, a synthetic wave is acquired by adding both components in a frequency domain and performing doubling reverse FFT, aperture credit, and overlap. In addition, separately, aperture credit and overlap are performed for reverse FFT, and both components may be added later. Moreover, about a harmonic component, the sine wave corresponding to each harmonic overtone is generated, and it may add with the non-harmonic component called for by reverse FFT, aperture credit, and overlap, and you may unite.

[0035] Drawing 9 is the functional block diagram showing more the configuration of the gestalt of 1 operation of the song synthesizer unit of this invention shown in said drawing 2 in a detail. In this drawing, the same sign is given to the same component as said drawing 2 . Moreover, in this example, in the phoneme (voice element) database 10, as for a harmonic component, the amplitude spectral envelope information for every frame shall be included, and, as for a non-harmonic component, the amplitude spectral envelope information and phase spectral envelope information for every frame shall be included. a words voice notation conversion means to by_which the words and a melody separation means of separating words data and melody data, and 32 change the words data from said words and melody separation means 31 into a voice notation (phoneme) train from the score data of a musical piece with which 31 should compound singing voice in drawing 9 -- it is -- the phoneme train from this words voice notation conversion means 32 -- said phoneme (a voice notation) -- base -- it is inputted into a piece conversion means 21. Moreover, the input of various control parameters, such as II Tempo which controls a performance, is enabled, the pitch information separated from score data with said words and melody separation means 31, dynamics information, such as a dynamic mark, and said control parameter are inputted into the pitch decision means 33, and the pitch of a song sound, dynamics, and II Tempo are determined. the base from said piece conversion means 21 of ****** -- information, such as piece information and a pitch from said pitch decision means, dynamics, and II Tempo, -- base -- the piece selection means 34 is supplied -- having -- the base from said voice element database (phoneme database) 10 with this most suitable piece selection means 34 of ** -- piece data are searched and outputted. the base which is completely in agreement with retrieval conditions at this time -- 1 which is similar when piece data are not memorized, or two or more base -- piece data are read.

[0036] said base -- the base outputted from the piece selection means 34 -- the data of the harmonic component of the piece data are supplied to the harmonic-component adjustment means 22. said base -- the base read by the piece selection means 34 -- when piece data are plurality, interpolation processing is performed so that it may agree on said retrieval conditions in the spectral envelope interpolation section 35 in this harmonic-component adjustment means 22, and the configuration of a spectral envelope is further deformed by the spectral envelope variant part 36 corresponding to said control parameter if needed. on the other hand -- said base -- the base outputted from the piece selection means 34 -- the data of the non-harmonic component of the piece data are inputted into the non-harmonic component adjustment means 23. The pitch information from said pitch decision means 33 is inputted into this non-harmonic component adjustment means 23, and as said drawing 3 was explained, compression or expanding processing of the frequency shaft according to a pitch is performed to the low-pass of a non-harmonic component. That is, with a band pass filter 37, the magnitude spectrum and phase spectrum of a non-harmonic component are trichotomized into low-pass, a mid-range, and a high region, and the compression or expanding of a frequency shaft corresponding to a pitch is performed in frequency axial compression and the expanding sections 38 and 39 about low-pass and a mid-range, respectively. The signal of a high region with which the signal of low-pass [on which compression or expanding processing of this frequency shaft was performed], and a mid-range, and such actuation are not made is supplied to the peak controller 40, and that peak value is adjusted so that the configuration of the spectral envelope of this non-harmonic component may be maintained.

[0037] The harmonic-component data from said harmonic-component adjustment means 22 and the non-harmonic component data from said non-harmonic component adjustment means 23 are inputted into the duration major key ready means 24. and the pronunciation time amount length determined using said melody information and said II Tempo information in this duration major key ready means 24 -- responding -- base -- a change of the time amount length of a piece is made. as mentioned above, base -- in shortening duration of piece data, when operating a frame on a curtailed schedule in time base compaction and the expanding section 43 and lengthening duration, loop-formation processing explained about said drawing 4 in the loop-formation section 42 is performed. the base which had duration length adjusted with said duration major key ready means 24 -- piece data perform level adjustment processing which was explained about said drawing 5 - drawing 8 with the level adjustment means 25 -- having -- base -- the piece connecting means 26 -- a harmonic component and each non-harmonic component -- it connects with time series.

[0038] said base -- the base connected by the piece connecting means 26 -- the harmonic component (spectral envelope information) of piece data is inputted into the harmonic-component generation means 27. To this harmonic-component generation means 27, the pitch information from said pitch decision means 33 is supplied, and the harmonic-overtone component corresponding to said pitch information according to said spectral envelope information is generated for it. Thereby, the actual harmonic component of the frame is obtained. and said base -- the amplitude spectral envelope information on the non-harmonic component from the piece connecting means 26 and phase spectral envelope information, and the magnitude spectrum of the harmonic component from said harmonic-component generation means 27 are compounded with an adder 28. And the signal of the frequency domain corresponding to each frame compounded in this way is changed into the wave signal of a time domain with the inverse Fourier transform means (reverse FFT means) 51, and the multiplication of the windowing function corresponding to frame length is carried out with the aperture credit means 52, and it compounds further, making the wave signal for every frame overlap with the overlap means 53. And the time amount wave signal compounded in this way is changed into an analog signal with the D/A conversion means 54, and it outputs from a loudspeaker 56 through amplifier 55.

[0039] Furthermore, drawing 10 is drawing showing an example of the hardware for operating the example shown in said drawing 9 . The central processing unit with which 61 controls actuation of this whole song synthesizer unit in this drawing (CPU), ROM, as for 62, various programs, a constant, etc. are remembered to be, RAM 63 remembers a work area and various data to be, The timer which 64 makes generate data memory, timer interruption predetermined in 65, etc., The words and the melody input section into which 66 inputs said score data, words data, etc. of the musical piece which should be performed, The control parameter input section into which 67 inputs each control parameter about a performance etc., The display as which 68 displays various information, the D/A converter from which 69 changes said compounded song data into an analog signal, and 70 are buses to which an amplifier and 71 connect a

loudspeaker to and 72 connects between said each component. Here, said phoneme database 10 is loaded on said ROM62 or RAM63, a song sound is compounded as mentioned above according to the data inputted from words, the melody input section 66, and the control parameter input section 67, and composite tone is outputted from a loudspeaker 71. The configuration shown in this drawing 10 is considered as the same configuration as the usual general purpose computer, and each above-mentioned function part of the song synthesizer unit of this invention can be realized also as an application program of a general purpose computer.

[0040] now, the base stored in said phoneme database 10 in the gestalt of operation mentioned above -- piece data were the magnitude spectrum and phase spectrum for every frame of the spectral envelope of every unit time amount (frame) of a harmonic component, and a non-harmonic component in SMS data and a typical example. and -- above -- a vowel etc. -- lengthening -- the base of a sound -- it was what can compound the song sound of high quality by memorizing piece data. However, especially, it lengthens and, in the case of a sound, there is a problem that the amount of data will become large since the harmonic component [in / it lengthens and / the time of day (frame) of all the sections of a sound] and a non-harmonic component are memorized. Since it should just have data for every frequency of the integral multiple of a basic pitch in the case of a harmonic component, a basic pitch needs to have 150Hz and the maximum frequency needs to have amplitude data (or also phase) about the frequency of 150 as 22025z, for example. On the other hand, in the case of a non-harmonic component, much more data are required, and it is necessary to have an amplitude spectral envelope and a phase spectral envelope about all frequencies. When the number of sampling points in one frame is made into 1024 points, the data of the amplitude and a phase are needed about the frequency of 1024. Since it is necessary to lengthen, to lengthen about a sound and to have data about all the frames in the sound section especially, the magnitude of data will become very big. Moreover, although it is desirable to prepare data for every various pitches in order to raise natural gender as mentioned above in addition to lengthening and preparing the data of the section of a sound for every phoneme, the amount of the data in a database will become still larger by this.

[0041] Then, the gestalt of other operations of this invention which can make size of said database very small is explained. With the gestalt of this operation, when creating said database 10, in case it lengthens and the data of the non-harmonic component of a sound are memorized, a spectral envelope whitening means is added. And he is trying to establish the spectral envelope generation means of a non-harmonic component in said non-harmonic component adjustment means at the time of composition. The need of lengthening and memorizing the spectral envelope according to an individual about the non-harmonic component of a sound by this is abolished, and reduction of the amount of data is enabled.

[0042] Drawing 11 is drawing in which lengthening and showing an example of the spectral envelope of the harmonic component in the case of a sound, and a non-harmonic component. As shown in this drawing, a vowel etc. lengthens, generally the configuration resembles the spectral envelope of a harmonic component, namely, the location of the spectral envelope of the non-harmonic component in the case of a sound of a crest or a trough corresponds about. Therefore, if some actuation (a gain adjustment, adjustment of an overall inclination, etc.) is performed to the spectral envelope of a harmonic component, a thing appropriate as a spectral envelope of a non-harmonic component can be obtained. Moreover, it lengthens and it is thought to a sound that the delicate fluctuation of each frequency componen in each frame within the object section is important, and it does not change so much even if the degree of this fluctuation changes a vowel. Then, the amplitude spectral envelope of a non-harmonic component is beforehand made flat in a certain form, and the effect of the tone of the vowel of a basis is removed (it whitens). Whitening considers as a flat spectrum at appearance. And at the time of composition, it asks for the spectral envelope of a non-harmonic component based on the configuration of the spectral envelope of a harmonic component, and if it applies to said spectral envelope which whitened, it can ask for the amplitude spectral envelope of a non-harmonic component. That is, only a spectral envelope is generated based on the spectral envelope of a harmonic component, and uses about a phase the thing from the first which lengthens and is contained in the non-harmonic component of a sound as it is. It becomes possible for a vowel which it whitened and which lengthens and is different based on sound data to lengthen, and to generate the non-harmonic component of sound data by doing in this way.

[0043] Drawing 12 is drawing for explaining creation processing of said phoneme database 10 in the gestalt of this operation of this invention, gives the same number to the same component as said drawing 1 , and decides to omit explanation. As shown in this drawing 12 , in the gestalt of this operation, it has a spectrum whitening means 80 to whiten the magnitude spectrum of the non-harmonic component which lengthens and is outputted from said section logging means 14 about a sound. Only the magnitude spectrum which lengthened and it whitened as a magnitude spectrum of the non-harmonic component of a sound by this is memorized, and only the phase spectrum will be memorized as a non-harmonic component of each piece data of **.

[0044] Drawing 13 is drawing showing the example of 1 configuration of said spectrum whitening means 80. As mentioned above, although it lengthens with this spectrum whitening means 80, it whitens the magnitude spectrum of the non-harmonic component of a sound and it considers as a flat thing at appearance, at this time, a spectrum is completely made into flatness (it has the same value on all frequencies) over not all the frames within the section, but actuation of carrying out near of the configuration of each frame evenly, with the delicate time fluctuation of each frequency left is needed. Then, as shown in drawing 13 , in the representation amplitude spectral envelope creation section 81, it asks for the typical amplitude spectral envelope within the section, and asks for the inverse number of each frequency component of the spectral envelope in the inverse number generation section 82 of a spectral envelope, and actuation of multiplying each frequency component of the spectral envelope of each frame by this in a filter 83 is performed. Here, in order to ask for the typical amplitude spectral envelope within said section, the average is taken for every frequency and it is good also as a typical spectral envelope using the average. Moreover, it is good also as a typical spectral envelope using the maximum of each frequency component within the section. Thereby, the magnitude spectrum which it whitened from said filter 83 is obtained. moreover, a phase spectrum -- as it is -- the base -- it memorizes to the non-harmonic component field of a piece.

[0045] Thus, although it lengthens and whitens the non-harmonic component of a sound, since it asks for a non-harmonic component using the spectral envelope of a harmonic component at the time of composition, the non-harmonic component which it whitened can be used common to all vowels, if it is a vowel. That is, it is enough, if it is a vowel and there is a non-harmonic component to lengthen and which whitened the sound. Of course, even if it has two or more whitening non-harmonic components, it does not interfere.

[0046] Drawing 14 is drawing for explaining the synthetic processing at the time of memorizing the magnitude spectrum which lengthened in this way and whitened about the non-harmonic component of a sound. In this drawing, the same number is given to the same component as said drawing 2 , and explanation is omitted. the base concerned read from said phoneme database 10 in the gestalt of

this operation as shown in this drawing -- a spectral envelope generation means 90 by which the non-harmonic component (white spectrum) of a piece is inputted is added to the preceding paragraph of said non-harmonic component adjustment means 24. As mentioned above, when the non-harmonic component which is lengthened from said phoneme database 10 and which whitened the sound is read, in the spectral envelope generation means 90, the amplitude spectral envelope of a non-harmonic component is calculated based on the spectral envelope of a harmonic component. For example, the component of the maximum frequency can consider how to determine that the spectral envelope of a non-harmonic component changes only the inclination of envelopment of a spectrum as what not changing. And this amplitude spectral envelope is inputted into said non-harmonic component adjustment means 24 with the phase spectral envelope of the non-harmonic component read to coincidence. The following processings are the same as that of the case where it is shown in said drawing 2 .

[0047] Thus, it lengthens, and the magnitude spectrum of the non-harmonic component which the part lengthened and it whitened only about the sound when the magnitude spectrum of the non-harmonic component of a sound was whitened and memorized is memorized, others lengthen, and it can avoid memorizing the magnitude spectrum of a non-harmonic component about a sound. in this case, the time of composition -- lengthening -- the base of a sound -- the phoneme nearest to [since there is no magnitude spectrum of a non-harmonic component in piece data] the phoneme to compound -- from the inside of a database -- choosing -- the -- what is necessary is to lengthen and just to create the magnitude spectrum of that non-harmonic component as mentioned above using the non-harmonic component of a sound Moreover, it lengthens, the group to whom the phoneme which divides into one or more groups the phoneme in which a sound is possible, and compounds it belongs lengthens, one of sound data is used, and you may make it generate the magnitude spectrum of a non-harmonic component as mentioned above.

[0048] In addition, when using the magnitude spectrum of a non-harmonic component for which it asked from the magnitude spectrum which it whitened as mentioned above, and the magnitude spectrum of a harmonic component Make it move so that it may be located near [corresponding to the pitch of the request to which the data near / corresponding to the pitch of former data / a harmonic overtone reproduce all or some of frequency shaft of a phase spectrum of the non-harmonic component] a harmonic overtone. That is, the phase data near a harmonic overtone become possible [considering as more natural composite tone] by making it use as phase data near a harmonic overtone also at the time of composition. Thus, according to the gestalt of this operation, it becomes possible to lengthen about all vowels, and for it to become unnecessary to memorize the non-harmonic component of a sound, and to reduce the amount of data in a database.

[0049] Furthermore, when defining the spectral envelope of a non-harmonic component by changing only the inclination of envelopment of this spectrum, whenever [of synthesized speech / husky] can be controlled by relating change of that inclination with "whenever husky]." That is, since it will become husky voice if there are many non-harmonic components in synthesized speech, and it will become charming voice if few, it will become charming voice if it will become husky voice if an inclination becomes sudden (the gain in 0Hz is large), and an inclination becomes gently-sloping (the gain in 0Hz is small). Then, whenever [of synthesized speech / husky] is controllable by controlling the inclination of the spectral envelope of a non-harmonic component by the parameter showing whenever husky] to be shown in drawing 15 .

[0050] Drawing 16 is drawing showing the example of a configuration of said spectral envelope generation means 90 at the time of enabling it to perform control whenever husky, and adds the property acquired by carrying out the multiplication of the inclination according to information to the spectral envelope of a harmonic component whenever [husky] it is supplied as a control parameter, and doing in this way with a filter 92 in the spectral envelope generation section 91 to the magnitude spectrum which whitened said non-harmonic component. And the phase spectral envelope of said non-harmonic component and the output of said filter 92 are outputted to said non-harmonic component controller 24 as data of a non-harmonic component.

[0051] In addition, the spectral envelope of a harmonic component may be modeled in a certain form, and whenever [husky] may be related with the parameter in it. For example, it may relate with whenever [husky] by changing either of the parameters when formulizing the spectral envelope of a harmonic component (parameter relevant to an inclination), and you may ask for the spectral envelope of a non-harmonic component. Moreover, whenever [husky] is good also as immobilization in time, and good also as adjustable. When it is made adjustable, the interesting effectiveness that voice becomes more husky and more husky can also be acquired while lengthening the phoneme.

[0052] Moreover, in order to enable it to only perform control whenever husky, it is not necessary to memorize the magnitude spectrum which the non-harmonic component whitened in the phoneme database 10 as mentioned above. The magnitude spectrum is memorized as it is like the piece. the gestalt of the first operation mentioned above -- like -- lengthening -- other base [non-harmonic component / of a sound] -- A flat spectrum is created by carrying out the multiplication of the inverse number to the magnitude spectrum of the non-harmonic component in quest of the magnitude spectrum which lengthens and represents within the sound section at the time of composition. What is necessary is to calculate the magnitude spectrum of a non-harmonic component according to the parameter which controls whenever [husky] based on the magnitude spectrum of a harmonic component, and just to let the spectrum obtained by carrying out multiplication to said flat spectrum be the magnitude spectrum of a non-harmonic component.

[0053]

[Effect of the Invention] According to the song synthesizer unit of this invention, the following effectiveness can be acquired as explained above.

- By use of an SMS technique, intelligibility is good and a synthetic song sound also with the lengthened natural part is obtained.
- By use of an SMS technique, even when delicate change of a vibrato or a pitch is performed, it does not become unnatural composite tone.
- the base in which the configuration of the spectral envelope of a voiced sound part (harmonic component) contains the optimal thing -- since it asks for a piece with selection or interpolation, change of the configuration of the spectral envelope by the pitch can also be coped with. Consequently, the tone which may be set in a broad pitch is obtained.
- In order to change a configuration with a detailed spectrum configuration about the non-harmonic component in the case of a voiced sound so that a desired pitch may be suited, even if it mixes a non-harmonic component and a harmonic component, there is no ******** squirrel ****** in ***************** by the noise.

- Since a phoneme lengthens and the die length of a part and the die length of a phoneme chain can be adjusted freely, a synthetic song sound can be obtained as desired II Tempo.

- About the connection parts of a phoneme and a phoneme, in order to perform level adjustment of smoothing, or its phoneme and phoneme, a noise does not occur at the time of connection.

- The compounded singing voice becomes a tone suitable for a desired pitch, is sung to the timing for which it asks, does not have a noise between connection units, either, and turns into singing voice of high quality.

[0054] Moreover, according to the song synthesizer unit of this application which lengthens, and whitens and memorizes the non-harmonic component of a sound, it becomes possible to both raise the effectiveness of database creation as if for size of a database to be made very small. Moreover, it becomes possible to offer the song synthesizer unit which can adjust the husky degree of synthesized speech easily.

[Translation done.]

CLAIMS

[Claim(s)]
[Claim 1] It has the phoneme database which memorized the data of a harmonic component, and the data of a non-harmonic component about the voice element which is the phoneme chain which is relation of a phoneme or two phonemes or more. By reading the voice element data corresponding to words from said phoneme database, and connecting A duration adjustment means to adjust the time amount length of the voice element data which are the song synthesizer unit which compounds a song sound, and were read from said phoneme database so that target II Tempo and how to sing might be suited, The song synthesizer unit characterized by having an adjustment means to adjust said harmonic component and said non-harmonic component of the voice element data read from said phoneme database so that the target pitch might be suited.
[Claim 2] the base which performs smoothing processing or level adjustment processing about a harmonic component and each non-harmonic component when connecting said voice element data -- the song synthesizer unit according to claim 1 characterized by having a piece level adjustment means.
[Claim 3] Claim 1 characterized by memorizing a pitch, dynamics, and two or more voice element data with which II Tempo differs about the same phoneme or a phoneme chain into said phoneme database, or a song synthesizer unit given in two.
[Claim 4] The song synthesizer unit according to claim 1 to 3 characterized by memorizing the voice element data which consist of the voice element data which consist of a phoneme chain from a vowel or a vowel to a consonant, voice element data which consist of a phoneme chain from a consonant to a consonant, and a phoneme chain from a vowel to a vowel from the voice element data which a vowel etc. lengthens and consist of a sound into said phoneme database, and a consonant.
[Claim 5] the data of said harmonic component, and the data of said non-harmonic component -- the base -- the song synthesizer unit according to claim 1 to 4 characterized by what is memorized as a data stream of the frequency domain corresponding to each frame of the frame train included at the section of a piece.
[Claim 6] Said duration adjustment means is a song synthesizer unit according to claim 5 characterized by to repeat 1 in the frame train included in a voice element, or two or more frames, or being what generates the frame train of desired time amount length by thinning out a frame.
[Claim 7] Said duration adjustment means is a song synthesizer unit according to claim 6 characterized by reversing the phase of the phase spectrum of the non-harmonic component when repeating the frame of a non-harmonic component and it goes back in time at the time of composition.
[Claim 8] The song synthesizer unit according to claim 5 characterized by having a harmonic-component generation means to change only a pitch into a desired pitch, maintaining the facies of the spectral envelope of the harmonic component contained in voice element data about a harmonic component at the time of song sound composition.
[Claim 9] It is the song synthesizer unit according to claim 5 characterized by having memorized the flat spectrum obtained by lengthening among the voice element data memorized in said phoneme database, and carrying out the multiplication of the inverse number of the spectrum which lengthens and represents the section of a sound to the magnitude spectrum of the non-harmonic component as a magnitude spectrum of a non-harmonic component about the voice element corresponding to a sound.
[Claim 10] It is the song synthesizer unit according to claim 9 characterized by obtaining the magnitude spectrum of a non-harmonic component by lengthening, calculating the magnitude spectrum of a non-harmonic component based on the magnitude spectrum of the harmonic component, and multiplying said flat spectrum by it about the non-harmonic component of a sound, at the time of song sound composition.
[Claim 11] They are claim 9 which the part in said phoneme database lengthens, uses said flat spectrum which the magnitude spectrum of the non-harmonic component is not memorized, but others lengthen about the voice element about a sound, and is memorized by the voice element of a sound, and is characterized by the thing [lengthening and compounding a sound], or a song synthesizer unit given in 10.
[Claim 12] The song synthesizer unit according to claim 10 characterized by controlling the gain in 0Hz of the magnitude spectrum of said non-harmonic component to calculate according to the parameter which controls whenever [husky] when calculating the magnitude spectrum of a non-harmonic component based on the magnitude spectrum of said harmonic component.
[Claim 13] At the time of song sound composition, lengthen, and to the magnitude spectrum of the non-harmonic component of a sound, carry out the multiplication of the inverse number of the representation [lengthen and] magnitude spectrum within the sound section, and a flat spectrum is created. The magnitude spectrum according to the parameter which lengthens and controls whenever [husky] based on the magnitude spectrum of the harmonic component of a sound is calculated. The song synthesizer unit according to claim 5 characterized for the magnitude spectrum obtained by multiplying by this magnitude spectrum and said created flat spectrum by the thing [lengthening and using it as a magnitude spectrum of the non-harmonic component of a sound].

[Translation done.]